

***pautolio***

a portfolio generator powered by AI

Karl Michael Jacinto (z5208448)

# Contents

[Contents 2](#_Toc48478669)

[The Problem 3](#_Toc48478670)

[Product Design and Use Cases 3](#_Toc48478671)

[What is *Pautolio*? 3](#_Toc48478672)

[Use Cases 4](#_Toc48478673)

[User wants to find the ideal combination of stocks that would maximise their returns given their risk profile 4](#_Toc48478674)

[User wants to act on their portfolio recommendation 4](#_Toc48478675)

[Interface 5](#_Toc48478676)

[Data Calculations and Storage Requirements 6](#_Toc48478677)

[Underlying Technology 7](#_Toc48478678)

[Input Features 7](#_Toc48478679)

[Model Design 8](#_Toc48478680)

[Portfolio Optimisation 8](#_Toc48478681)

[Sentiment Analysis 9](#_Toc48478682)

[Incorporating Sentiments into Portfolio Optimisation 11](#_Toc48478683)

[Model Implementation 13](#_Toc48478684)

[Model Limitations 16](#_Toc48478685)

[Appendix 17](#_Toc48478686)

[Extra plots 17](#_Toc48478687)

[Stock correlation heatmap 17](#_Toc48478688)

[Simulated portfolio returns 17](#_Toc48478689)

[Optimal Risky Portfolio and Capital Market Line 18](#_Toc48478690)

[Asset plots 18](#_Toc48478691)

[Scaled sentiment histogram 22](#_Toc48478692)

[Bibliography 22](#_Toc48478693)

# The Problem

2020 has been a difficult year for investors. COVID-19 and the subsequent economic downturn across the world has led to deep instability in the financial markets, with some commentators even identifying a “disconnect between Wall Street and Main Street” [[1]](#_Bibliography). The pandemic has upended the state of the world, altering how we work, socialise, and spend our downtime. The world of investing is no different – for instance, the Australian Financial Review recently published some recommendations for where to invest “in a post-pandemic world” [[2]](#_Bibliography). Indeed, with the world economy in a state of limbo, **people are unsure of what to invest in**.

Traditional investing has primarily been driven by the frameworks of fundamental analysis, which involves analysing the financial statements of companies, and technical analysis, which harnesses historical prices and market statistics. However, in the past few years, the growth of machine and deep learning has led to an explosion in the use of quantitative analysis – the use of mathematical and statistical tools - to study investments. The reading of news articles to gather insights on the economy, industries, and companies has always been a part of fundamental analysis, however, thanks to deep learning, there are now widely-used **methods available to *quantify* the impact of news** articles and headlines on share prices and investor sentiment. With the growth of social media, information is travelling faster and more widely than ever before, and hence stocks are more prone to changes in sentiment.

The investment landscape in recent times has also witnessed growing support for passive, rather than active, investing. A study by Morningstar concluded that passive and exchange-traded funds have delivered better returns than actively managed funds in the last five years [[3]](#_Bibliography). However, more and more young people are becoming educated in active investing thanks to the growth of robo-advisors and forums such as the Personal Finance Subreddit [[5]](#_Bibliography). But what about in between? There seems to be a lack of investment solutions that combine the returns and simplicity of a passive fund, whilst **giving some control back to the investor**. This is where *Pautolio* fits into the investment landscape.

# Product Design and Use Cases

## What is *Pautolio*?

*Pautolio* is an AI-driven investment manager that aims to strike a balance between providing an optimal portfolio recommendation for investors and giving them some control over aspects of investing such as risk tolerance. Investors can enter their risk tolerance and their level of capital before an algorithm automatically recommends an optimal portfolio that updates daily. A user can then quickly purchase an approximate portfolio through the app. More details on *Pautolio*’s use cases can be found under the subsection “Use Cases”.

The target demographic for *Pautolio* consists of:

* Young active investors who want to simplify the asset selection process
* Young passive investors who want to move to investing in individual stocks.

*Pautolio* is available as both a web app and a smartphone app. Because the app targets younger people, the smartphone app will be developed first.

## Use Cases

### User wants to find the ideal combination of stocks that would maximise their returns given their risk profile

Upon starting the app for the first time, users will be prompted to register using their email account. This process only requires the user’s name, their email, and their phone number (as a backup measure).

Once registered, users will be given the choice of a “Simple” portfolio optimisation or “Custom” portfolio optimisation. In the Simple portfolio, the user can choose from a number of “groups” of stocks to analyse in detail:

* *Top 10 Market Capitalisation*: these stocks consist of the ten largest ASX 200 firms by market capitalisation. They are also the ten stocks analysed in the code.
* *Financial Firms*: this group will consist of the big banks and financial services firms such as Commonwealth Bank and IAG.
* *Tech Tycoons*: this group will consist of major IT and telecommunications firms such as Xero and Telstra.
* *Ethical Enterprise*: this group consists of firms that focus on social impact, clean energy, and sustainability. These include The a2 Milk Company and Sonic Healthcare. With 76% of Generation Z and 71% of Millennials concerned about sustainability and the environment [[6]](#_Bibliography), having an ethical investment option would appeal to *Paultolio*’s main target demographic.

In the Custom portfolio, the user can choose which stocks in the ASX 200 will be used in the calculation.

After choosing their portfolio, users will enter how much capital they have. After, they will undergo a quick series of psychometric tests and personality questions to determine their level of risk aversion, a metric that can easily be changed later.

The app will then calculate the optimal portfolio and return a visual dashboard that displays approximately how much money should be invested in each stock, as well as how much money should be invested in a risk-free asset (a government bond).

### User wants to act on their portfolio recommendation

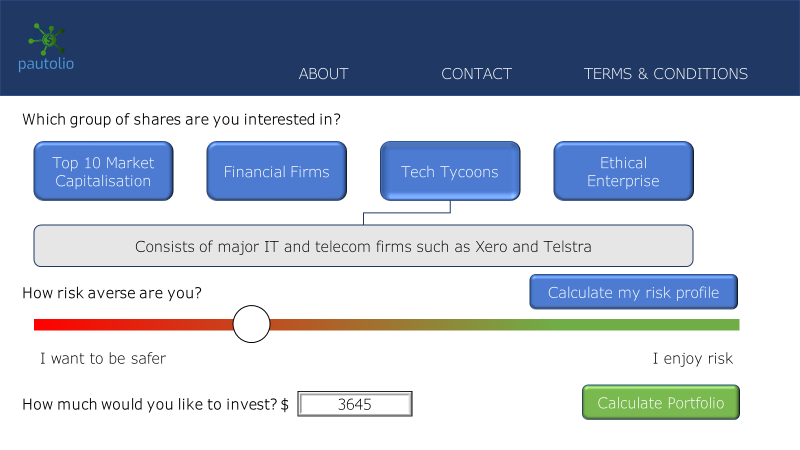
If a user wants to actually purchase the recommended portfolio, they can do so through the app. *Pautolio* will partner with brokerage firms such as CommSec to allow users to quickly place an order. If they want to purchase the portfolio, they will be redirected to the CommSec app/website. Once the user enters their personal details, CommSec will automatically fill out an order and ask for the user’s permission to place the order.

Despite the convenience of this feature, it does have some limitations:

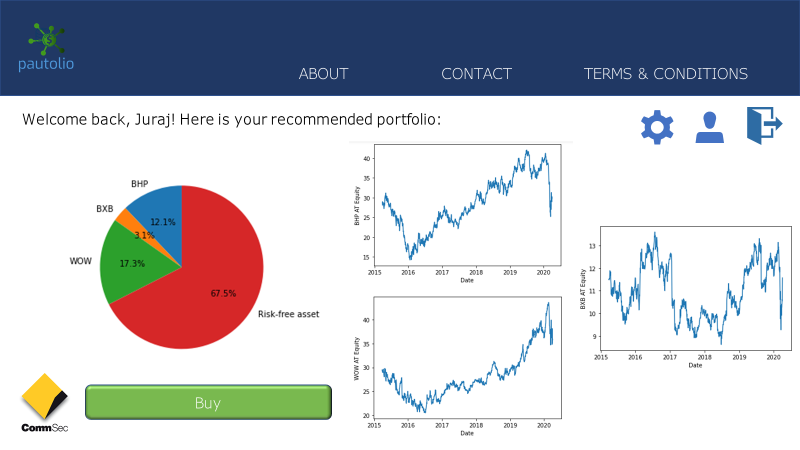
* Although *Pautolio* returns a portfolio recommendation in exact dollar amounts, in practice, the actual portfolio will be different as shares can only be purchased in integer amounts. This is particularly problematic for users with low capital amounts.
* Furthermore, *Pautolio* does not consider brokerage fees, which investors do have to pay every time they place an order on CommSec.

## Interface

It is well known in the UX space that younger generations place much more emphasis on visual design than their older counterparts – one source even states that ‘the new generation loves with the eyes, not with the ears’ [[4]](#_Bibliography). *Pautolio* strives for a user interface that is clean and simple.

The Simple portfolio selection process screen consists of just three inputs: the stock group the user selected, their risk tolerance, and their level of capital. The three inputs are also located on the one page, which minimises the number of clicks a user has to make. 

Once the portfolio has been calculated, the app displays a dashboard displaying the optimal portfolio, as mentioned beforehand. The dashboard takes up about two-thirds of the page, and consists of a pie chart of the ideal portfolio, and a line graph of the recent performance of each stock. At the bottom-left of the page is a green button that directs users to buy the portfolio through CommSec.



Careful attention was placed on the colour palette used for the app. *Pautolio* uses predominantly navy blue, cornflower blue, and olive green as accents on a white background. Blue typically symbolises trust and technology, whereas the green symbolises money and wealth – these values are key to a FinTech app.

## Data Calculations and Storage Requirements

Data to be calculated:

* Risk profile (upon user request)
* Sentiment value of news (updated at the end of each trading day)
* Portfolio weights and adjusted weights (updated at the end of each trading day)

Data to be stored:

* Risk profile
* Past news and share price data, with new data simply appended onto the existing datasets at the end of each trading day.

# Underlying Technology

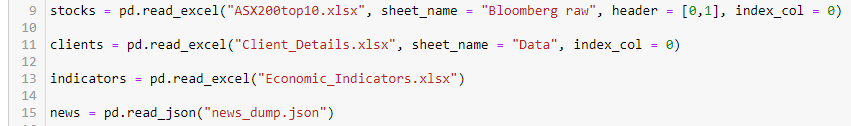
The following section explains how the portfolio optimisation in *Pautolio* works.

## Input Features/Feature Engineering

To calculate the optimal portfolio, multiple pieces of data are required:

* Daily adjusted closing share price data on ten different stocks from 30th March 2015 in the form of a csv file. These ten stocks are the 10 largest ASX-listed companies by market capitalisation. Although this data was provided as part of the task, data on other shares can be easily obtained using the Yahoo Finance API.
* A database of IMDB movie reviews to train the neural network for sentiment analytics. This database is built into TensorFlow (the Python library used for the sentiment analysis model).
* A JSON dataset of news headlines (with dates) pertaining to each stock. This dataset was provided as part of the task, but additional headlines can be obtained via webscraping or downloading via news website APIs.
* A csv dataset on the Consumer Sentiment Index, which measures consumer confidence in Australia. This dataset was provided as part of the task as part of a larger dataset of economic indicators that can be downloaded from sources such as FactSet.
* A float between 1 to 10, representing the user’s level of risk aversion. In the complete financial product, this is an input by the user. However, in this task, a list of clients and their risk profiles are provided.

In the code, the datasets are imported into the Python script at different times. For instance, the share prices, news headlines, client data, and economic indicators are imported near the beginning of the code:



On the other hand, the IMDB movie reviews were imported halfway through the code, at the start of the sentiment analysis:



After the equity prices, news, clients, and economic data was imported, it had to be cleaned and manipulated so that it was in a usable format. This involved:

* Removing all other columns except the “PX\_LAST” column for each stock; this column represents the closing price.
* Converting the closing prices for each stock into daily returns.
* Extracting only the Consumer Sentiment Index from the economic indicators dataset.
* Upsampling the Consumer Sentiment Index data, i.e. converting it from monthly to daily data.
* Converting the training and testing datasets for the RNN into padded batches.
* Cleaning the “Date/Time” field in the news dataset and converting it from a DateTime value to a Date value.

## Model Design

The overall model used to put together a portfolio recommendation consists of three parts:

* Non-adjusted portfolio recommendations
* Conducting sentiment analytics on the news dataset
* Adjusting the portfolio weights based on the sentiments

### Portfolio Optimisation

The non-adjusted portfolio recommendations were constructed using Markowitz (modern) portfolio theory, which states that investors are risk averse, and hence investors can construct a portfolio that maximises returns for a given level of risk [[7]](#_Bibliography).

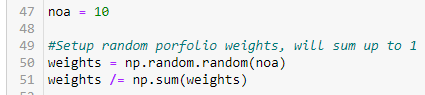
According to Markovitz, there exists what is known as an “Optimal Risky Portfolio” or “Optimal Tangent Portfolio”: a combination of assets that maximises the Sharpe Ratio. The Sharpe Ratio is a measure of the performance of an investment adjusted for risk:

where:

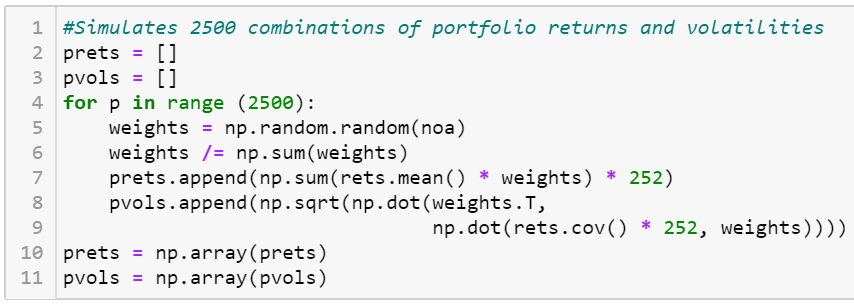
* E(rp) is the expected portfolio return
* rf is the risk-free rate, which in the code is assumed to be 1% per annum.
* σp is the standard deviation of the portfolio return

This portfolio, which is denoted as P\*, lies on the “Efficient Frontier”, which is a line connecting the portfolio with the highest return for each level of risk. P\* is the portfolio of risky assets that *all* investors, in theory, should invest in regardless of risk tolerance. Risk profiles come into play when determining what proportion of an investor’s capital should be invested in the Optimal Risky Portfolio. The more risk-averse an individual is, the greater the proportion of their funds that should be invested in the risk-free asset (a government bond). This proportion, denoted by y\*, is given by the following formula:

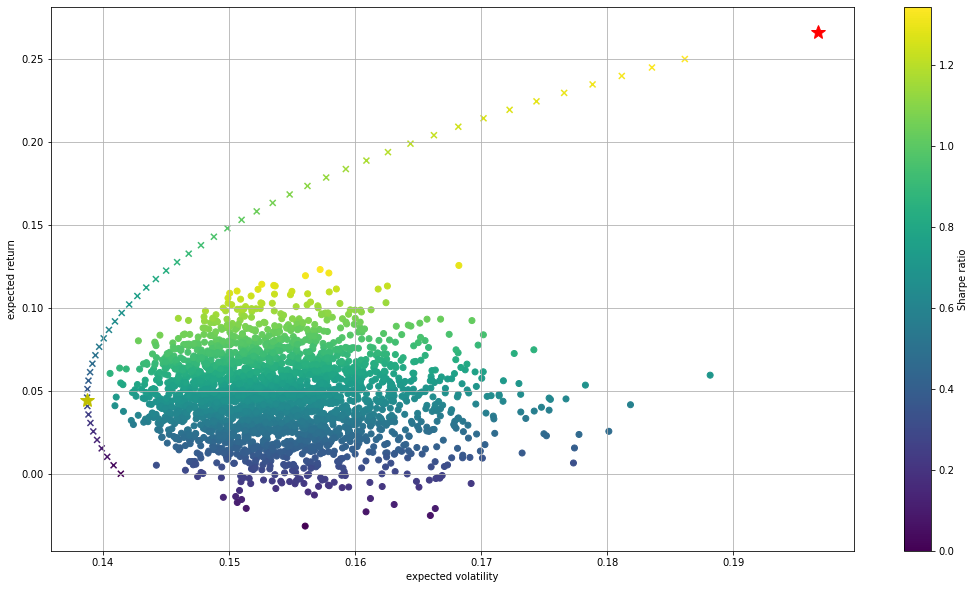
The model begins by randomly generating ten random weights, one corresponding to each asset and collectively adding to 1:



We then calculate the overall expected portfolio return and standard deviation, and use this to simulate 2,500 combinations of returns and standard deviations.



Given these combinations, we can then construct the Efficient Frontier, and calculate the Optimal Risky Portfolio:

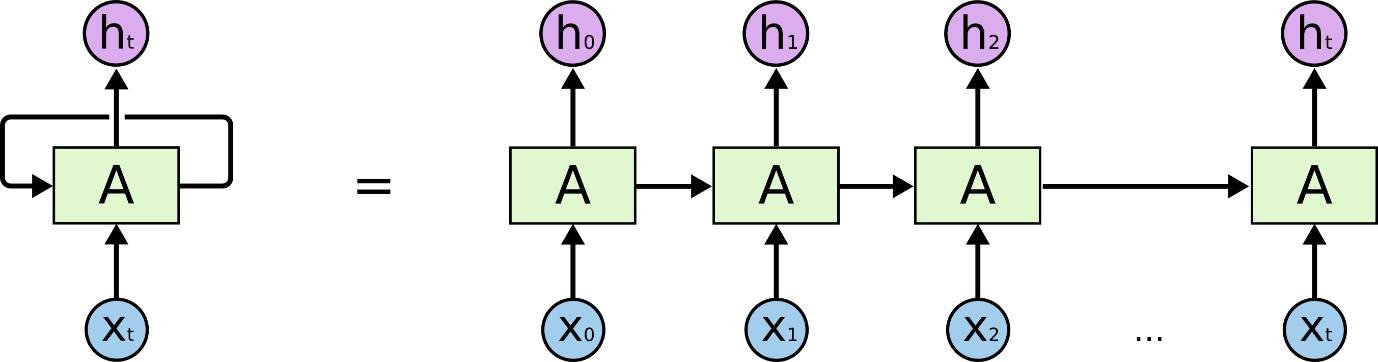


From this, we then use Python’s SciPy library to optimise the portfolio weights and calculate the risky share y\*. In this model, we assume no brokerage fees and no short selling, and hence if a portfolio weight would be negative, it is instead set to zero with the other weights then adjusted accordingly.

### Sentiment Analysis

After the base optimised portfolio was created, the data was ready to undergo sentiment analytics. The purpose of sentiment analytics is to adjust the weights of the Optimal Tangent Portfolio to take into account investor attitudes towards certain companies.

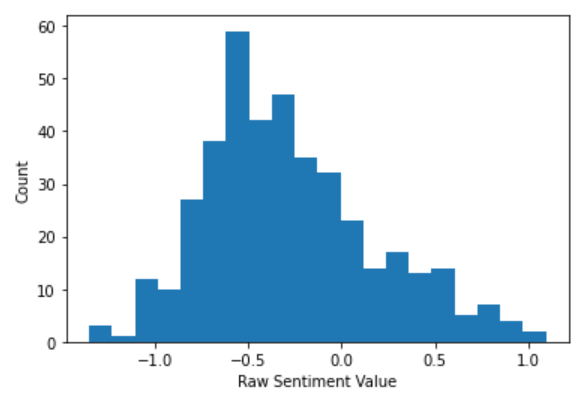
To predict the sentiment of the news articles, a recurrent neural network (RNN) was trained on the IMDB movie review database. An RNN is a type of neural network in which each node *i* takes in an input x­i and combines it with the previous node’s output hi-1 to produce a new output hi­. This allows the network to recognise sequential characteristics of data to predict the next outcome [[8]](#_Bibliography), and hence is useful for evaluating sentences.



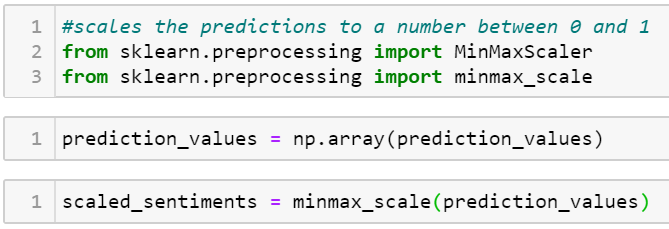
The RNN was trained using the following model assumptions:

* Buffer size of 10,000
* Batch size of 64
* Layer structure:
  + One embedding layer with 64 nodes
  + Two bi-directional layers:
    - The first layer consists of 64 nodes and returns the full sequence after each node
    - The second layer consists of 32 nodes and returns only the last output for each node
  + One dense layer with 64 nodes using the “relu” activation function
  + One dropout layer with rate 0.5. This layer prevents the model from overfitting.
  + One dense layer with one node.

After the model was trained, it was used to predict the sentiment of each of the news headlines. In theory, the sentiments should all lie between 0 and 1, however, the model predicted a negative number for most news headlines.

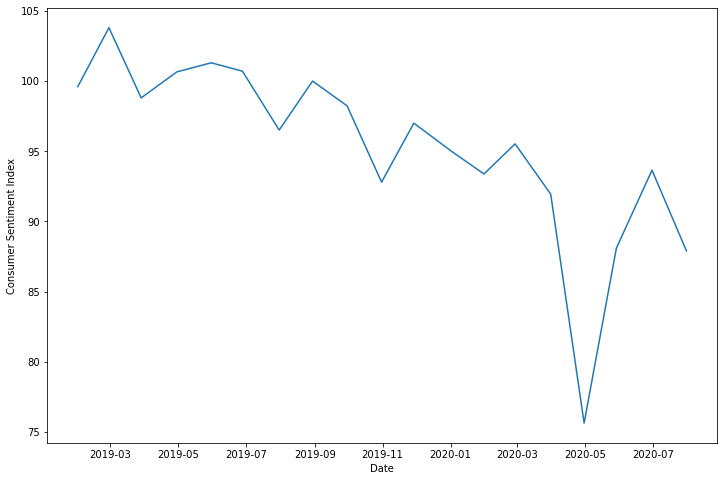


To gain a clearer understanding of whether certain sentiments were positive or negative, the **raw sentiment values were scaled to between 0 and 1** using Scikit-Learn’s Min-Max scaler.



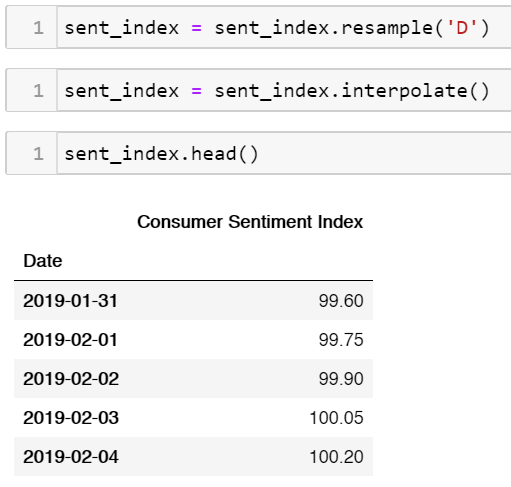
### Incorporating Sentiments into Portfolio Optimisation

Differences in sentiment values may be due to positive or negative public perception toward the companies. However, **sentiments may also fluctuate due to macroeconomic events**. For instance, it is more likely that sentiment values would be down during a recession than during a boom. Indeed, a graph of the Consumer Sentiment Index indicates a gradual fall in sentiment from its peak in early 2019, before a monumental plunge in May of 2020 due to the COVID-19 pandemic.

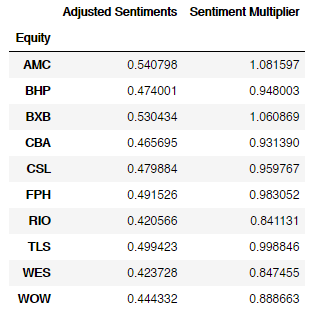


Consequently, it was important to **scale the individual sentiments** to reflect fluctuations in overall consumer sentiment. This was achieved by obtaining the mean of the Consumer Sentiment Index values (95.375) and applying the following formula:

Because the adjusted sentiment values require the exact date each headline was written, we need to obtain daily data for Consumer Sentiment Index. This was achieved byusing **linear interpolation** to gather daily values from the monthly data, as mentioned in the “Input Features” section.



The **mean sentiment for each stock** was obtained by aggregating the sentiment dataset by company. Each adjusted sentiment was then divided by 0.5 (“neutral” sentiment) to give a “**Sentiment Multiplier**”: a factor that each weight in the optimised portfolio would need to be multiplied by to account for the sentiments.



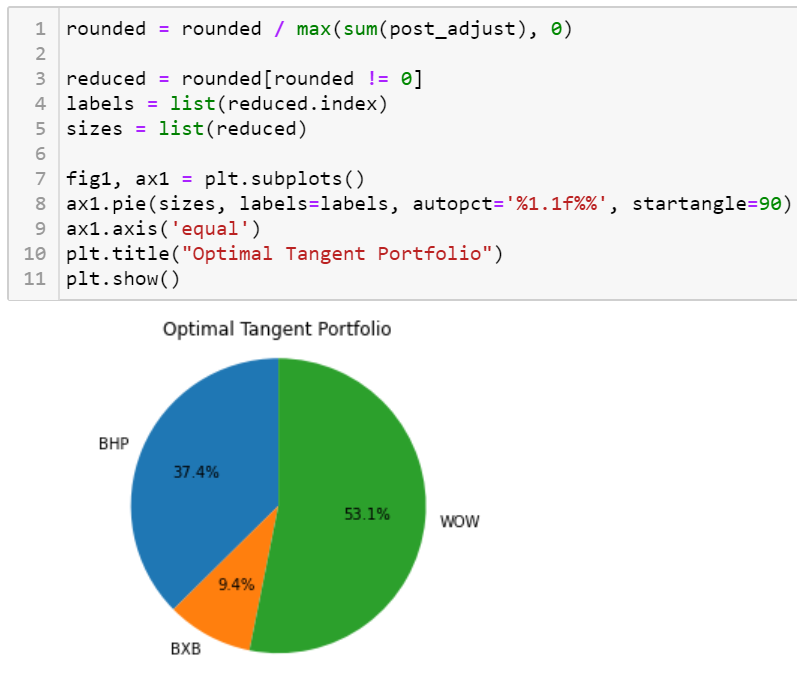
The Sentiment Multipliers would be multiplied by the optimised portfolio weights to give a final, adjusted set of portfolio weights.

## Model Implementation

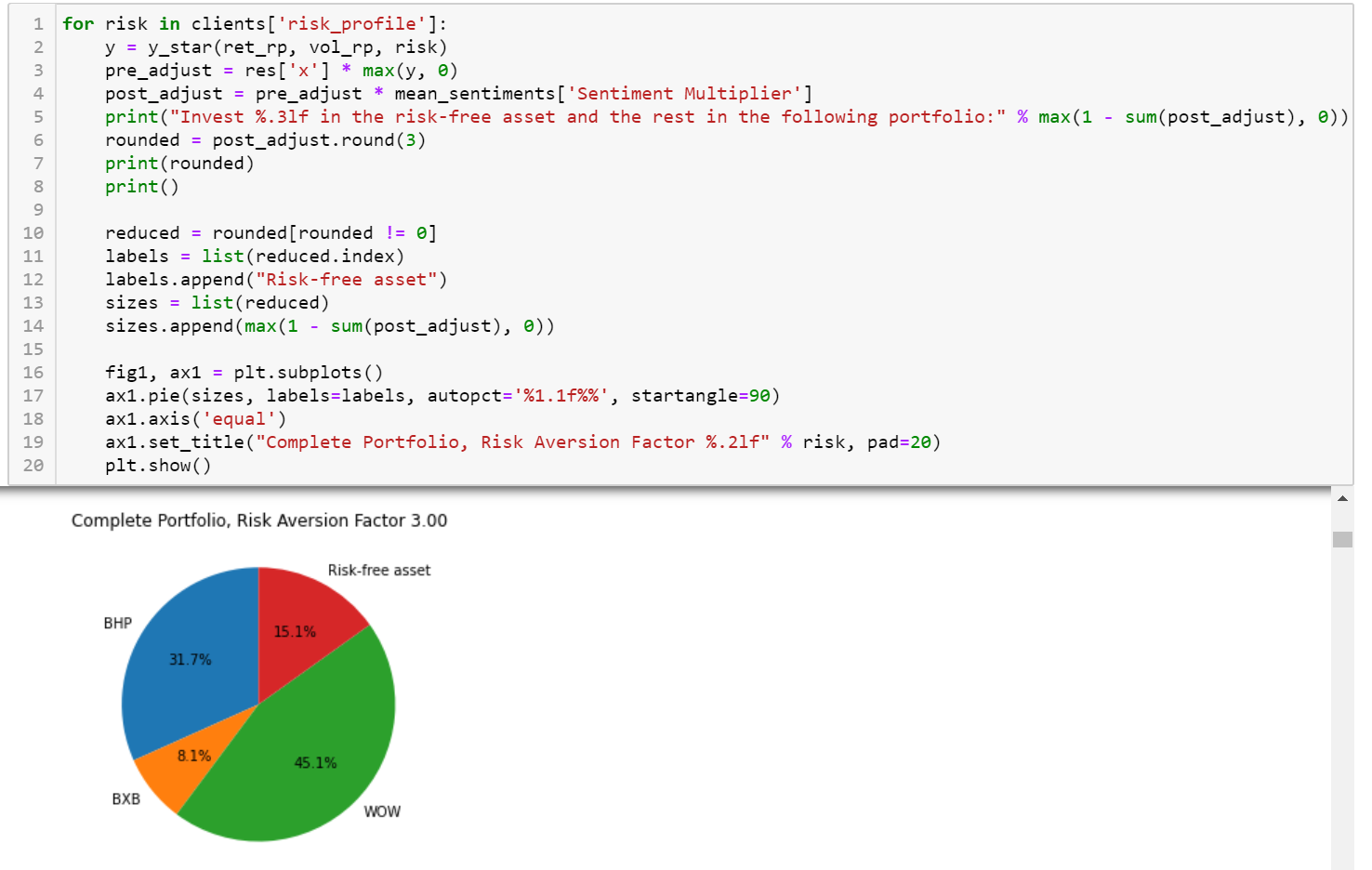
To assist clients in making effective and informed investment decision-making, the code produces a number of different outputs.

*Pautolio* will produce a summary of the efficient frontier and Optimal Risky Portfolio in the form of two charts: a scatter plot of the 2,500 portfolio simulations and the efficient frontier, and a pie chart of P\*:

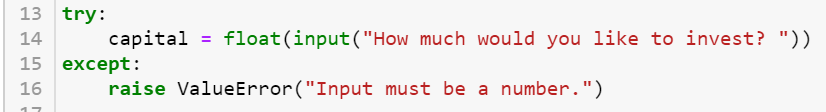




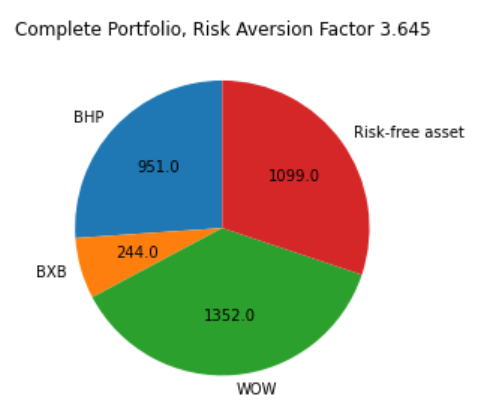
From this, the program takes in the user’s risk appetite as a number from 1 to 10 and calculates a recommended portfolio. The portfolio weights are then adjusted to account for the sentiment values described in the “Model Design” section, before the final portfolio is then displayed to the user in the form of a pie chart:



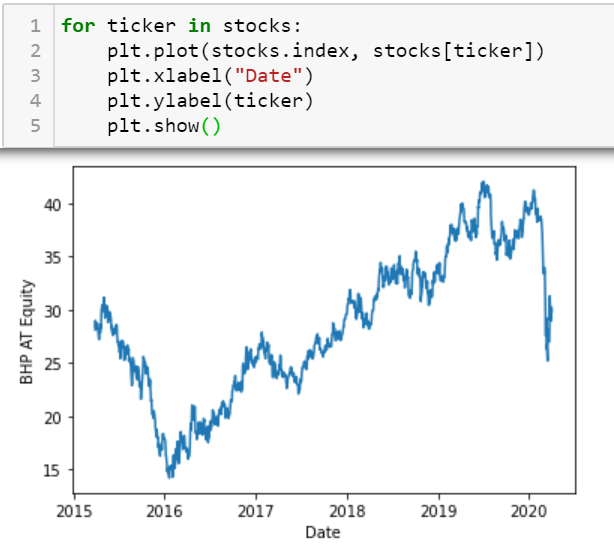
*Pautolio* can also be extended to account for the amount of capital a user has.



If the user inputs the total funds they have, the program can produce a pie chart with currency units:



The user will also receive charts displaying the performance of each of the stocks.



As mentioned in the Product Design section, the app could be extended through a partnership with CommSec (or another brokerage firm). CommSec will help build an extension to the app that connects Pautolio with CommSec, allowing users to invest in the recommended portfolio quickly and easily.

## Model Limitations

Despite the model’s ability to successfully produce a portfolio recommendation, there are some limitations that inhibit the model’s effectiveness in a practical, real-world situation:

* As mentioned in the “Model Design” section, the model assumes **no brokerage fees** and **no short selling**.
* Does **not consider the user’s current portfolio position**, but this is alleviated due to the no-brokerage-fee assumption. If there were brokerage fees, moving to the optimal portfolio may be more expensive than holding on to their current position in the short- to medium-term, especially if they have very low capital.
* **Assumes that the risk-free rate is 1%**. In reality, the risk-free rate, which is determined as the interest rate on government bonds, depends on the term of the bond.
* Model **does not consider other user demographics** such as age, or other economic indicators, both of which are provided in the given datasets.
* **Slow to train** RNN (but fast to predict on new headlines)
* Model **does not give higher weight to more recent headlines**, which are more likely to affect the current stock price.

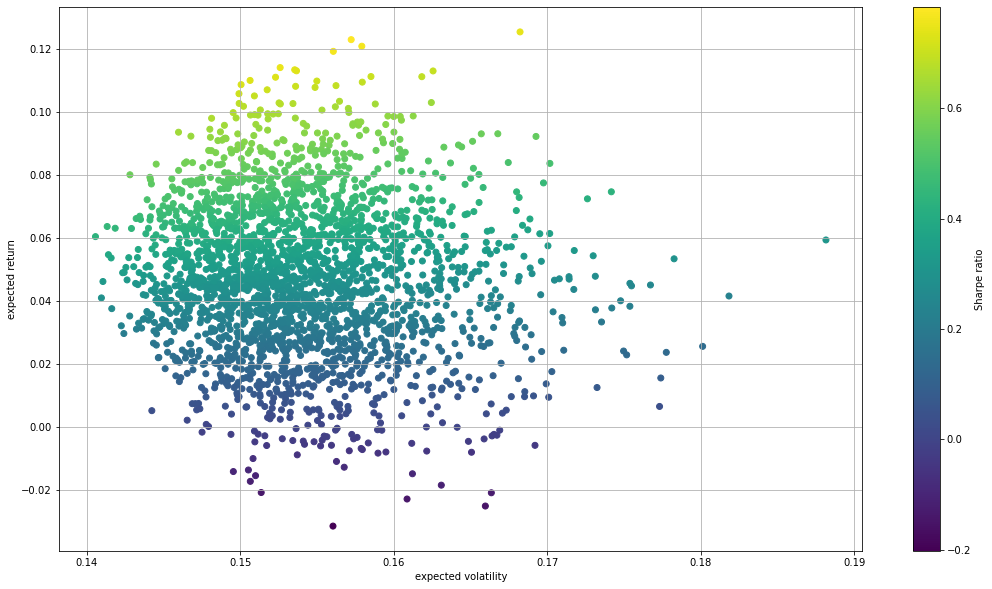
# Appendix

## Extra plots

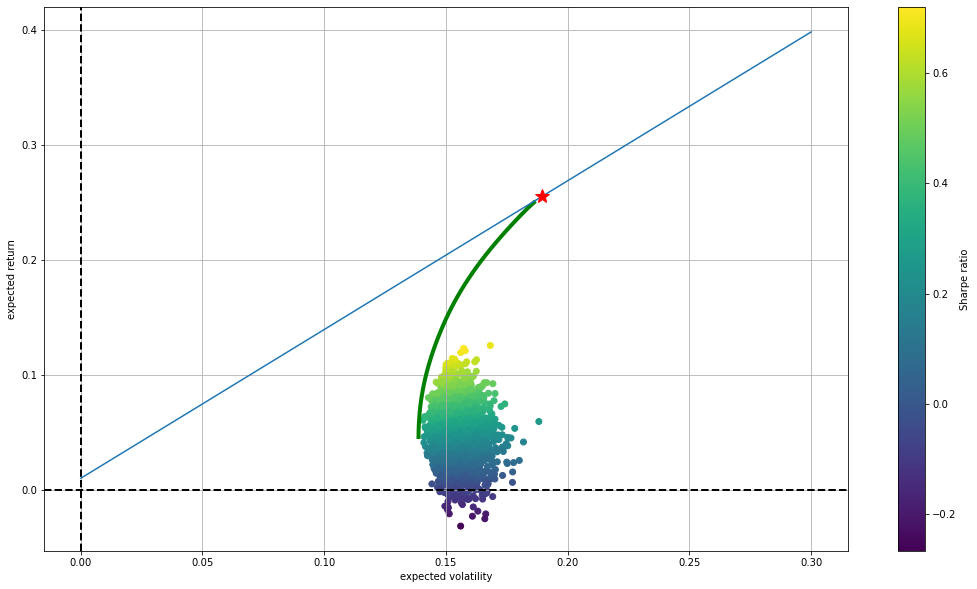
### Stock correlation heatmap



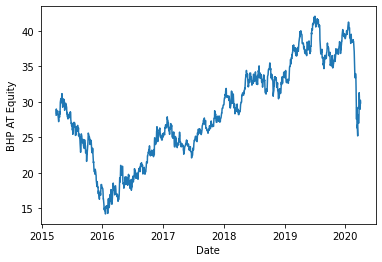
### Simulated portfolio returns

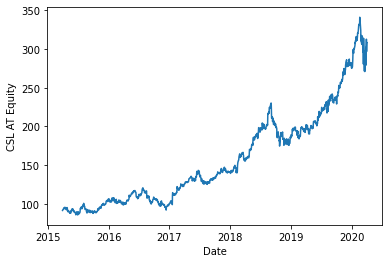
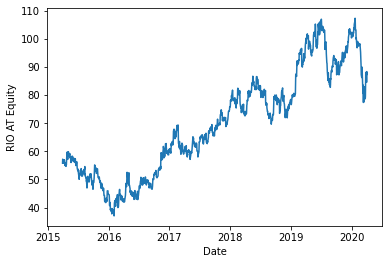
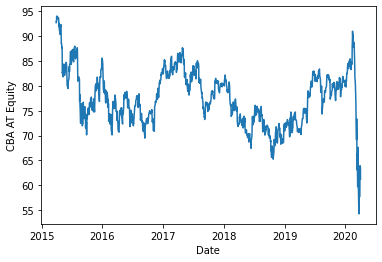
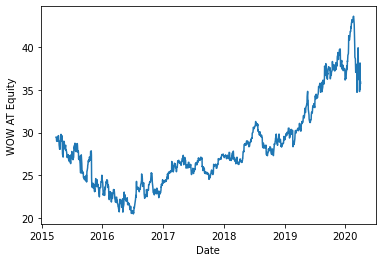
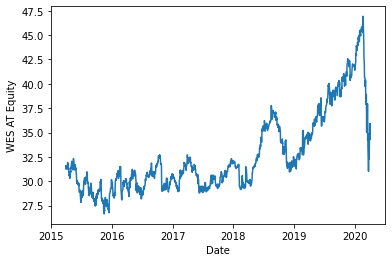
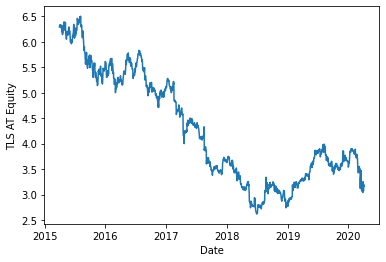
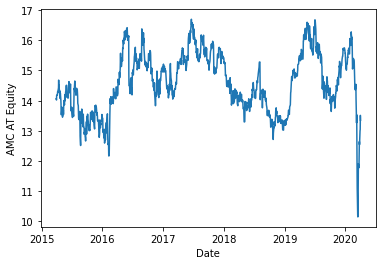
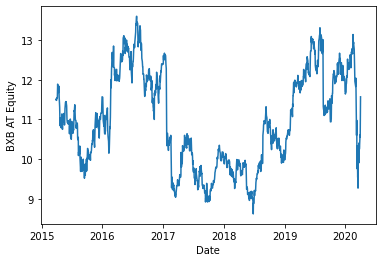
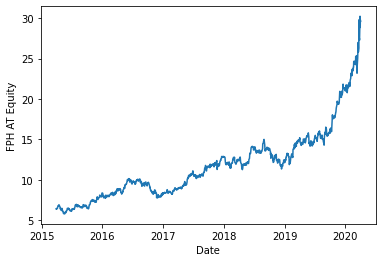


### Optimal Risky Portfolio and Capital Market Line

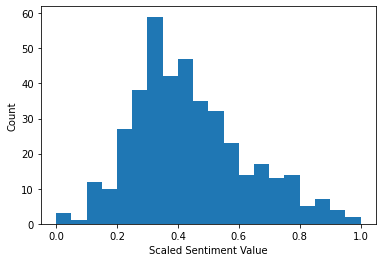


### Asset plots



### Scaled sentiment histogram



## Bibliography

1. <https://www.bloomberg.com/opinion/articles/2020-05-04/coronavirus-mend-disconnect-between-wall-street-and-main-street>
2. <https://www.afr.com/policy/economy/where-to-invest-post-covid-19-20200813-p55ldv>
3. <https://www.forbes.com/sites/jamesphillipps/2020/08/13/passives-lead-the-way-as-active-funds-fail-to-deliver/#4a4192eaf4e7>
4. <https://www.uxdesignagency.com/blog/banking-should-meet-customer-expectations>
5. <https://www.morningstar.com.au/learn/article/young-investors-react-to-covid-19/201526>
6. <https://www.finder.com.au/sustainability-facts-and-stats>
7. <https://www.investopedia.com/terms/m/modernportfoliotheory.asp>
8. <https://developers.google.com/machine-learning/glossary/#recurrent_neural_network>